Q

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/904,253	07/12/2001	Srinivas Bangalore	1999-0779	2050

7590 06/22/2005

Samuel H. Dworetsky
AT&T CORP.
P.O. Box 4110
Middletown, NJ 07748-4110

EXAMINER	
CHAWAN, VIJAY B	
ART UNIT	**PAPER NUMBER**
2654	

DATE MAILED: 06/22/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/904,253	BANGALORE ET AL.	
	Examiner	**Art Unit**	
	Vijay B. Chawan	2654	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☐ Responsive to communication(s) filed on _____.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) 1-53 is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) 1-53 is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on 16 October 2001 is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

Claim Objections

1. Claims 2, 6, 11, 12, 16, 19, 29 are objected to because of the following informalities: Above claims have language in the claims that make the claims ambiguous and indefinite. For example, in claim 2, it is not clear when "a first system that inputs... and **that** outputs...", it is not clear what the applicant means by "**that.**" In claim 6, line 3, "... outputs a recognition results relating finite-state machine", there appears to be some claim language missing. In claim 11, it is not clear what the applicant means by "that other mode recognition system". It is not clear from the claims that claim 11, depends upon do not clearly set forth what are the other mode(s). In claim 19, it is not clear how the second mode relates to the first mode to a meaning of a combination of the first and second modes, and also, by what is meant by " a possible meaning". Does the applicant mean likelihood model? Errors such as these are contained throughout the claim language exists and should be corrected to make clear the claimed invention. Appropriate correction is required.

Claim Rejections - 35 USC § 101

2. 35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

3. Claims 1-53 are rejected under 35 U.S.C. 101 because the claimed invention is directed to non-statutory subject matter.

Claims 1-53, define non-statutory processes because they merely manipulate an abstract idea without a claimed limitation to a practical application. The disclosed invention has an application in the technological arts (viz. Speech Recognition and/or Gesture recognition), however, the claimed system simply manipulates an abstract idea without a claimed limitation to the practical application and does not have any pre or post-computer activity. See MPEP 2106, Section IV.

A review of the application 09/904,253 shows the disclosed invention describing in the specification as a general purpose computer and/or blocks showing an implementation of a mathematical algorithm, not a physical component or a circuit.

Applicant should note that that claimed limitation directed to, for example, receiving an utterance to be processed would be considered to be statutory subject matter.

Claim Rejections - 35 USC § 102

4. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

5. Claims 1-53 are rejected under 35 U.S.C. 102(b) as being anticipated by Sharma et al., ("Toward Multimodal Human-Computer Interface", Proceedings of the IEEE, vol.86, Issue 5, may1998, pages 853-869).

As per claim 1, Sharma et al., teaches a finite-state multi-modal recognition system that generates a multimodal meaning based on an utterance comprising a plurality of associated modes, the system comprising:

a plurality of finite-state mode recognition systems, each finite-state mode recognition system usable to recognize ones of the associated modes, each finite-state mode recognition system outputting at least one recognition lattice for each associated mode (Fig.3); and,

an n-tape finite-state device that inputs n-1 recognition lattices from the plurality of finite-state mode recognition subsystems and outputs the multimodal meaning based on the n-1 recognition lattices (Fig.3, section III, pages 856-858).

As per claim 2, Sharma et al., teaches a finite-state multimodal recognition system that generates a multimodal meaning based on an utterance comprising a pair of associated modes, the system comprising:

a pair of finite-state mode recognition systems, each finite-state mode recognition system usable to recognize one of the associated modes, each finite-state mode recognition system outputting at least one recognition lattice for each associated mode (Fig.3, section III, pages 856-858); and,

a multimodal recognition system that inputs a recognition lattice from each of the pair of mode recognition systems and outputs the multimodal meaning for the pair of associated modes based on the plurality of recognition results comprising:

a first system that inputs the pair of recognition lattices and outputs a combined recognition finite-state transducer (Fig.3, section III, pages 856-858);

a second system that inputs the combined recognition finite-state transducer and outputs a combined recognition finite-state machine (Fig.3, section III, pages 856-858); and,

a third system that inputs the combined recognition finite-state machine and a multimodal meaning grammar and outputs the multimodal meaning (Fig.3, section III, pages 856-858).

As per claim 3, Sharma et al., teaches a finite-state multimodal recognition system that generates a multimodal meaning based on an utterance comprising a pair of associated modes, the system comprising:

a plurality of mode recognition subsystems, each mode recognition subsystem usable to recognize ones of the associated modes, each mode recognition subsystem outputting at least one recognition result for each associated mode (Fig.3, section III, pages 856-858); and,

a multimodal recognition subsystem that inputs recognition results from each of the plurality of mode recognition subsystems and outputs the multimodal recognition for the plurality of associated modes based on the plurality of recognition results, wherein each of the plurality of mode recognition subsystems and the multimodal recognition

subsystems and the multimodal recognition subsystem includes at least one finite-state machine having at least on tape (Fig.3, section III, pages 856-858).

As per claim 4, Sharma et al., teaches the multimodal recognition system of claim 3, wherein the multimodal recognition subsystem comprises a first subsystem that inputs the recognition results from at least one of the plurality of mode recognition subsystems and that generates a first finite-state transducer that relates the input information results from each of the at least one mode recognition subsystems to a recognition model of at least one other mode recognition subsystem (Section C, pages 861-863).

As per claim 5, Sharma et al., teaches the multimodal recognition system of claim 4, wherein the multimodal recognition subsystem further comprises a second subsystem that inputs the first finite-state transducer and the recognition results from the at least one other mode recognition subsystem and that generates a second finite-state transducer based on the recognition results from the at least one other mode recognition subsystem and the first finite-state transducer (section VI, pages 864-866).

As per claim 6, Sharma et al., teaches the multimodal recognition system of claim 5, wherein the multimodal recognition subsystem further comprises a third subsystem that inputs the second finite-state transducer and outputs a recognition result relating to said finite-state machine (section VI, pages 864-866).

As per claim 7, Sharma et al., teaches the multimodal recognition system of claim 6, wherein the multimodal recognition subsystem further comprises a third finite state transducer, and a multimodal recognizer that inputs the first finite-state machine

and outputs the multimodal recognition based on the first finite-state machine and the third finite-state transducer (section VI, pages 864-866).

As per claim 8, Sharma et al., teaches the multimodal recognition system of claim 7, wherein the multimodal recognition is multimodal meaning (section VI, pages 864-866).

As per claim 9, Sharma et al., teaches the multimodal recognition system of claim 4, wherein the first subsystem comprises at least one finite-state transducer, each second finite-state transducer relating the recognition results of one of the plurality of mode recognition systems to the recognition model of the at least one other mode recognition subsystem, and a second subsystem that generates the first finite-state transducer based on the input recognition results from the at least one mode recognition subsystem and the at least one second finite-state transducer (section VI, pages 864-866).

As per claim 10, Sharma et al., teaches the multimodal recognition system of claim 9, wherein the first subsystem further comprises a third subsystem that generates at least one projection of the first finite-state transducer, each projection output to a corresponding one of the at least one other mode recognition subsystem (section VI, pages 864-866).

As per claim 11, Sharma et al., teaches the multimodal recognition system of claim 10, wherein each projection output to a corresponding one of the at least one other mode recognition subsystem is usable as a recognition model by said mode recognition subsystem (section VI, pages 864-866).

As per claim 12, Sharma et al., teaches the multimodal recognition system of claim 10, wherein each other mode recognition subsystem inputs the corresponding projection as a recognition model usable to recognize the at least one associated mode that is recognized by that other mode recognition subsystem (section VI, pages 864-866).

As per claim 13, Sharma et al., teaches the multimodal recognition system of claim 3, wherein the plurality of mode recognition subsystems comprise at least two of a gesture recognition subsystem, a speech recognition subsystem, a pen input recognition subsystem, a computer vision recognition subsystem, a haptic recognition subsystem, a gaze recognition subsystem, and a body motion recognition system (section VI, pages 864-866).

As per claim 14, Sharma et al., teaches the multimodal recognition system of claim 13, wherein the plurality of mode recognition subsystems include at least a first mode recognition subsystem that inputs a first one of the plurality of different modes and outputs a first mode recognition lattice as the recognition result of the first mode recognition subsystem and a second mode subsystem that inputs a second one of the plurality of different modes and outputs a second mode recognition lattice as the recognition result of the second mode recognition subsystem (section VI, pages 864-866).

As per claim 15, Sharma et al., teaches the multimodal recognition system of claim 14, wherein the multimodal recognition subsystem comprises a first subsystem that inputs the first mode recognition lattice from the first mode recognition subsystem,

and that generates a first finite-state transducer that relates the first mode recognition lattice to a recognition model of the second mode recognition subsystem (section VI, pages 864-866).

As per claim 16, Sharma et al., teaches the multimodal recognition system of claim 15, wherein the multimodal recognition subsystem further comprises a second subsystem that inputs the first finite-state transducer and the second mode recognition lattice from the second mode recognition subsystem, and that generates a second finite-state transducer based on the second mode recognition lattice from the second mode recognition subsystem and the first finite-state transducer (section VI, pages 864-866).

As per claim 17, Sharma et al., teaches the multimodal recognition system of claim 16, wherein the multimodal recognition subsystem further comprises a third subsystem that inputs the second finite-state transducer and outputs a first finite-state transducer (section VI, pages 864-866).

As per claim 18, Sharma et al., teaches the multimodal recognition system of claim 17, wherein the multimodal recognition subsystem further comprises a third finite-state transducer, and a multimodal recognizer that inputs the first finite-state machine and outputs the multimodal recognition based on the first finite-state machine and the third finite-state transducer (section VI, pages 864-866).

As per claim 19, Sharma et al., teaches the multimodal recognition system of claim 18, wherein the third finite-state transducer relates the first mode and the second mode to a meaning of a combination of the first and second modes, and, the multimodal

recognizer comprises a meaning subsystem that inputs the first finite-state machine and outputs, as the multimodal recognition, a possible meaning lattice based on the first finite-state machine and the third finite-state transducer (section VI, pages 864-866).

As per claim 20, Sharma et al., teaches the multimodal recognition system of claim 15, wherein the first subsystem comprises a second finite-state transducer that relates the first mode recognition lattice from the first mode recognition system to the recognition model of the second mode recognition subsystem, and, a second subsystem that generates the first finite-state transducer based on the input first mode recognition lattice and the second finite-state transducer (section VI, pages 864-866).

As per claim 21, Sharma et al., teaches the multimodal recognition system of claim 20, wherein the first subsystem further comprises a third subsystem that generates a projection of the first finite-state transducer (section VI, pages 864-866).

As per claim 22, Sharma et al., teaches the multimodal recognition system of claim 21, wherein the projection is output to the second mode recognition subsystem and is usable as a recognition model by the second mode recognition subsystem (section VI, pages 864-866).

As per claim 23, Sharma et al., teaches the multimodal recognition system of 21, wherein the second mode recognition subsystem inputs the projection as a recognition model usable to recognize at least the second mode input by the second mode recognition subsystem (section VI, pages 864-866).

As per claim 24, Sharma et al., teaches the multimodal recognition system of claim 3, wherein the plurality of mode recognition subsystems includes at least a

gesture recognition subsystem that inputs a gesture mode and outputs a gesture recognition lattice as the recognition result of the gesture recognition subsystem and a speech recognition subsystem that inputs at least one speech mode and outputs a word sequences lattice as the recognition result of the speech recognition subsystem (section VI, pages 864-866).

As per claim 25, Sharma et al., teaches the multimodal recognition system of claim 24, wherein the multimodal recognition subsystem comprises a first subsystem that inputs the gesture recognition lattice from the gesture recognition subsystem and that generates a first finite-state transducer that relates the gesture recognition lattice to a recognition model of the speech recognition subsystem (section VI, pages 864-866).

As per claim 26, Sharma et al., teaches the multimodal recognition system of claim 25, wherein the multimodal recognition subsystem further comprises a second subsystem that inputs the first finite-state transducer and the word sequences lattice from the speech recognition subsystem and that generates a second finite-state transducer based on the word sequences lattice from the speech recognition subsystem and the first finite-state transducer (section VI, pages 864-866).

As per claim 27, Sharma et al., teaches the multimodal recognition system of claim 26, wherein the multimodal recognition subsystem further comprises a third subsystem that inputs the second finite-state transducer and outputs a first finite-state machine (section VI, pages 864-866).

As per claim 28, Sharma et al., teaches the multimodal recognition system of 27, wherein the multimodal recognition subsystem further comprises a third finite-state

transducer, and a multimodal recognizer that inputs the first finite-state machine and outputs the multimodal recognition based on the first finite-state machine and the third finite-state transducer (section VI, pages 864-866).

As per claim 29, Sharma et al., teaches the multimodal recognition system of claim 28, wherein the third finite state transducer relates the gesture mode and the speech mode to a meaning of a combination of the gesture and speech modes, and the multimodal recognizer comprises a meaning subsystem that inputs the first finite-state machine and outputs as the multimodal recognition a possible meaning lattice based on the first finite state machine and the third finite state transducer (section VI, pages 864-866).

As per claim 30, Sharma et al., teaches the multimodal recognition system of claim 25, wherein the first subsystem comprises a second transducer that relates the gesture recognition lattice from the gesture recognition systems to the recognition model of the speech recognition subsystem, and a second subsystem that generates the first finite-state transducer based on the input gesture recognition lattice and the second finite-state transducer (section VI, pages 864-866).

As per claim 31, Sharma et al., teaches the multimodal recognition system of claim 30, wherein the first subsystem further comprises a third subsystem that generates a projection of the first finite-state transducer (section VI, pages 864-866).

As per claim 32, Sharma et al., teaches the multimodal recognition system of claim 31, wherein the projection is output to the speech recognition subsystem and is

usable as a recognition model by the speech recognition subsystem (section VI, pages 864-866).

As per claim 33, Sharma et al., teaches the multimodal recognition system of claim 31, wherein the speech recognition subsystem inputs the projection as a recognition model usable to recognize the at least one speech mode input by the speech recognition subsystem (section VI, pages 864-866).

As per claim 34, Sharma et al., teaches the multimodal recognition system of claim 25, wherein the first subsystem comprises a second finite-state transducer that relates the gesture recognition lattice from the gesture recognition system to a language model of the speech recognition system as the recognition model of the speech recognition subsystem, and a second subsystem that generates as the first finite-state transducer, a gesture/language model finite-state transducer based on the input gesture recognition lattice and the second finite-state transducer (section VI, pages 864-866).

As per claim 35, Sharma et al., teaches the multimodal recognition system of claim 34, wherein the first subsystem further comprises a third subsystem that generates a projection of the gesture/language model finite-state transducer (section VI, pages 864-866).

As per claim 36, Sharma et al., teaches the multimodal recognition system of claim 35, wherein the projection is output to the speech recognition subsystem and is usable as a language model by the speech recognition subsystem (section VI, pages 864-866).

As per claim 37, Sharma et al., teaches the multimodal recognition system of claim 35, wherein the speech recognition subsystem inputs the projection as a language model usable to recognize the at least one speech mode input by the speech recognition subsystem (section VI, pages 864-866).

As per claim 38, Sharma et al., teaches the multimodal recognition system of claim 25, wherein the recognition model is one of a grammar model or a language model (section VI, pages 864-866).

As per claim 39, Sharma et al., teaches the multimodal recognition system of claim 24, wherein the gesture recognition subsystem comprises a gesture feature extraction subsystem that inputs the gesture mode and outputs a gesture feature lattice and a gesture recognition subsystem that inputs gesture feature lattice, and outputs the gesture recognition lattice (section VI, pages 864-866).

As per claim 40, Sharma et al., teaches the multimodal recognition system of claim 24, wherein the speech recognition system comprises, a speech processing subsystem that inputs a speech signal and outputs a feature vector lattice, a phonetic recognition subsystem that inputs the feature vector lattice and an acoustic model lattice and outputs a phone lattice, a word recognition subsystem that inputs the phone lattice and a lexicon lattice and outputs a word lattice, and a speech mode recognition subsystem that inputs the word lattice and a recognition model and outputs a word sequences lattice (Feature-level, pages 61-863, section VI, pages 864-866).

As per claim 41, Sharma et al., teaches the multimodal recognition system of claim 40, wherein the recognition model is input from the multimodal recognition subsystem (Fig.3).

As per claim 42, Sharma et al., teaches the multimodal recognition system of claim 3, further comprising a plurality of mode input devices, at least two of the plurality of mode input devices inputting different modes (Fig.3).

As per claim 43, Sharma et al., teaches the multimodal recognition system of claim 42, wherein the plurality of mode input devices comprises at least two of a gesture input device, a speech input device, a pen input device, a computer vision device, a haptic input device, a gaze input device, and a body motion input device (Fig.3).

As per claim 44, Sharma et al., teaches the multimodal recognition system of claim 43, wherein at least two of the plurality of input devices are combined into a single multimodal input device (Fig.3).

Claims 45-53 are method claims to be implemented on the system claims 3-44, and are similar in scope and content and are rejected under similar rationale.

Conclusion

6. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Bennett et al., (6,665,640) teach interactive speech based learning/training system formulating search queries based on natural language parsing of recognized user queries.

Brand (6,735,566) teaches generating realistic facial animation from speech.

Chen et al., ("Gesture-Speech Based HMI for a Rehabilitation Robot", Proceedings of the Southeastcon '96, "Bringing Together Education, Science and Technology"., 11-14, April 1996, pages 29-36), teach recognition of the spoken input to be used to supplant the need for general purpose object recognition between different objects and to perform the critical function of disambiguation.

Roy et al., ("Word Learning in a Multimodal Environment", proceedings of the 1998 IEEE Conference on Acoustics, Speech, and Signal Processing, 1998, ICASSP'98, 12-15 May 1998, vol.6, pages 3761-3764), teach building trainable interfaces which let the user teach the interface which words and gestures she wants to use and what the words and gestures mean. These trainable interfaces can also be used for gestures and other non-speech modalities.

Salem et al., ("Current Trends In Multimodal Input Recognition", IEE Colloquium on Virtual Reality Personal Mobile and Practical Applications – 98/454 28 Oct 1998, pages 3/1-3/6), teach different pattern recognition techniques currently available in the area of speech recognition, facial gesture recognition, body tracking and hand gesture recognition

Kettebekov et al., ("Toward Multimodal Interpretation in a Natural Speech/Gesture Interface", Proceedings 1999 International Conference on information and Intelligence

Systems, pages 328-335), teach the design of a speech/gesture interface in the context of a set of spatial tasks defined on a computerized campus map.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Vijay B. Chawan whose telephone number is (571) 272-7601. The examiner can normally be reached on Monday Through Friday 6:30-3:00.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (571) 272-7602. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Vijay B Chawan

Vijay B. Chawan
Primary Examiner
Art Unit 2654

vbc
6/16/05

VIJAY CHAWAN
PRIMARY EXAMINER